

Semiconductor device and method of operating such a device

The invention relates to a semiconductor device comprising a semiconductor body which is provided with a field effect transistor at a surface and which comprises strongly doped source and drain zones and a channel region extending between the source zone and the drain zone, with a gate electrode being present which overlaps the channel

5 region upon perpendicular projection thereon, wherein the source zone, the drain zone and the gate electrode are connected at the surface to a metal source contact, a drain contact and a gate electrode contact, respectively, and wherein a further metal strip is positioned between the gate electrode contact and the drain contact, which metal strip is insulated from the semiconductor body, is locally electrically connected to the source strip and forms a shield

10 between the gate electrode and the drain contact. Such a device may be used to advantage as a (high) power amplifier, in particular for a transmitter in the high-frequency range of wireless communication. The presence of the shield between the gate electrode contact and the drain contact leads to a significant capacity reduction between the drain contact and the gate electrode, which in turn leads to a significant power amplification increase at higher

15 frequencies.

Such a device, which may be constructed both in discrete and in integrated form, is known from US patent US 6,069,386, which was published on 30 May 2000. The

20 known device comprises a lateral metal oxide semiconductor-type transistor. The transistor is formed in a weakly doped p-type epitaxial layer on a strongly doped p-type substrate. The channel is defined in a p-type zone which is implanted into the epitaxial layer in a self-aligned manner with respect to the polycrystalline silicon gate electrode. The source and drain zones may have an interdigitated structure, the number of digits being chosen, for

25 example, in dependence on the maximum electric current to be accommodated. The drain extension serves to increase the breakdown voltage between source and drain, as is generally known. Since the metal source contact is not arranged above but beside the gate electrode, the gate electrode can be provided with a metal connection over its entire length, so that the gate resistance is determined by the layer resistance of the metal, and accordingly it can be

kept very low through the use of a suitably conductive metal. The contacts are in the form of parallel metal strips arranged side by side. Since the further metal strip and the metal contact of the gate electrode can be manufactured in the same metal layer as the source and drain contacts, relatively few process steps are necessary for the manufacture thereof.

5 One drawback of the known device is the fact that it does not have a linear characteristic, in particular at very high frequencies. As a result, the maximum usable power is limited.

10 Accordingly, the object of the present invention is to provide a device of the kind referred to in the introduction, which is highly linear, also at high frequencies, and which is capable of providing a high power level.

According to the invention, a semiconductor device of the kind referred to in the introduction is characterized in that the electrical connection between the further metal strip and the source contact comprises a capacitor, and in that the further metal strip is provided with a connecting contact for applying an external voltage to the further metal strip. The invention is in the first place based on the recognition that the provision of a capacitor in the connection between the further metal strip and the contact source enables the application to the further metal strip of a voltage which is different from the voltage of the source region.

15 20 25 30 In order to enable the application of such a voltage to the further metal strip, said metal strip is according to the invention provided with a connecting contact. The invention is furthermore based on the recognition that the linearity of the device can be significantly improved by applying such a voltage to the further metal strip. This is based on the recognition that the device according to the invention may be considered to comprise two field effect transistors, namely a first enhancement-type transistor, which is associated with the gate electrode, and a second depletion-type transistor, whose further metal strip forms the gate electrode, as it were. Since the former transistor operates in the so-termed common source configuration in the device according to the invention, and the latter transistor operates in the so-termed common gate configuration, the two transistors operate in a cascade (= series connected) configuration and exhibit opposite phase distortion which enables phase extinction and thus less distortion. Said further contact region makes it possible, through the application of an external voltage, to set the power level at which the aforesaid extinction takes place. This enables optimization of the linearity in dependence on the selected use of the device.

The capacitor may be externally connected to the device. A device which is ready for use is normally provided with an insulating envelope of plastic material, from which a number of terminal pins, four terminal pins in an advantageous embodiment of a device according to the invention, extend. The capacitor may be bonded to the enveloped 5 device, for example, and be connected to the terminal pins for the source region and for the further metal strip.

In an advantageous embodiment, the capacitor is integrated in the semiconductor body and is positioned within the active region beside the transistor. Such a device is easy and inexpensive to manufacture. Preferably, the source contact, the drain 10 contact, the gate electrode contact, the further metal strip and the connecting contact thereof and an electrode of the capacitor are formed of two separate metal layers arranged one above the other, which are separated from each other by a further insulating layer.

Such a two-metal layer process makes it possible to connect the further metal strip to the (one) electrode of the capacitor at several positions, which connections cross the 15 gate electrode contact. The use of a large number of such connections is advantageous with a view to lowering the impedance for the operating frequency between the further metal strip and the capacitor electrode. Another important aspect of such a two-metal layer process is the fact that two adjacent gate electrode digits may be connected via the upper layer of the two metal layers. Each digit may be connected to a conductor track in the upper metal layer by 20 means of a conducting via. In that case a conducting digit will be present - in the upper metal layer - between two adjacent gate electrodes, to which digits the conductor tracks of the two adjacent gate electrodes that lead to a via are connected. The conducting digit is connected to the contact region for the gate electrode, crossing the underlying source contact, which is interrupted at the location of said crossing so as to reduce the capacity. Preferably, two 25 adjacent gate electrodes are connected to the conducting digit at several positions in this manner. For technical reasons, each gate electrode is preferably widened at the location of said connection. The advantage of such a connection of the gate electrode to the gate electrode connecting region is that the connecting resistance of the gate electrode may be low, whereas the width thereof may nevertheless be relatively small. Preferably, in order to 30 reduce said resistance, the polycrystalline silicon of the gate electrode, too, has been partially converted into a silicide, such as a titanium silicide.

In another variant, the other electrode of the capacitor is formed by the semiconductor body, which comprises a strongly doped substrate, on which a more weakly

doped epitaxial layer is present. The substrate is connected to a metal layer, which also forms a connection for the source region.

In an advantageous variant, the two electrodes of the capacitor form part of the metal layers, and the lower electrode of said two electrodes is electrically connected to the 5 semiconductor body, which comprises a strongly doped region at that location. In this way a voltage-independent operation of the capacitor can readily be realized, which is advantageous. In this case silicon nitride may be used as the dielectric of the capacitor. Another advantage of such a capacitor is the fact that a higher capacity per unit area can be achieved, thus enabling a reduction of the surface area that is taken up by a capacitor having 10 a given capacity. The strongly doped region comprises, for example, the substrate and a more strongly doped (by means of a local diffusion) portion of a (more weakly doped) epitaxial layer present on the substrate.

The capacitance of the capacitor partially depends on the desired frequency. Preferably it ranges between 10 pF and 1 nF at an operating frequency ranging between 100 15 MHz and 3 GHz. Since a device according to the invention can be used with higher power levels than the prior art device, a device according to the invention is preferably provided with a heat sink, which preferably comprises copper and/or tungsten-copper parts.

The invention may be used to advantage in (n-channel or p-channel) transistors, in which the channel is formed by a surface region of the semiconductor body, 20 and in which the channel is separated from the gate electrode by means of an electrically insulating layer. A preferred embodiment of a device according to the invention is characterized in that the transistor is of the lateral DMOS type, in which D stands for "double diffused". The invention is also suitable for use in other transistors, however, such as MESFET's, and in particular HEMT's, which are based on a semiconductor body of a III-V 25 material, such as GaAs or GaN. Furthermore it is advantageous if a heat sink is present on the side of the semiconductor body remote from the surface.

Preferably, a minimum spacing is used between the further metal strip and the gate electrode contact. This makes it possible to influence other characteristics of the transistor as well via the voltage on the strip. Preferably, this takes place in an independent 30 manner, with another metal strip – which is likewise insulated from the semiconductor body – being present between the further metal strip and the gate electrode, which other strip is connected in the same manner as in the prior art device, for example. Said other strip may furthermore be connected in the same manner as the further metal strip in the device according to the invention. In that case the other transistor characteristics can be influenced

through the application of a different external voltage to the other metal strip, independently of the voltage being applied to the further metal strip.

The invention furthermore comprises a method of operating a device according to the invention, in which a voltage is applied to the contact region of the further metal strip during operation of the device. The applied voltage may be dynamically controlled, that is, the voltage that is applied in the higher power region is different from the voltage that is applied in the lower power region. According to another possibility, a gradually changing, power-dependence voltage is advantageously applied. In this variant, use is preferably made of the capacitor that is entirely formed in the two metal layers. Such a capacitor can be more easily connected to a circuit that is integrated in the device, by means of which a specific voltage function can be applied.

The above and further aspects of the invention will now be explained in more detail by means of an embodiment. In the drawing:

Fig. 1 is a top plan view of a semiconductor device according to the invention; Fig. 2 is a top plan view of an enlarged portion, indicated II in Fig. 1, of the device of Fig. 1;

Fig. 3 is a sectional view of this device along the line III-III;
Fig. 4 is a sectional view of this device along the line IV-IV;
Fig. 5 is a sectional view of a variant of this device along the line III-III;
Fig. 6 shows magnitudes illustrating the amplitude and phase linearity of the device of Fig. 1 as a function of the input power;

Fig. 7 shows a distortion factor and the amplification of the device of Fig. 1 as a function of the average power;

Fig. 8 shows two the distortion factors and the efficiency of the device as a function of the average power.

It is noted that the drawing is merely diagrammatic and not drawn true to scale. It is further noted that the metallization pattern is mainly shown in Fig. 1 and Fig. 2. Parts which lie at a lower level are not shown in Fig. 1 and Fig. 2, but only in the sectional views of Fig. 3, Fig. 4 and Fig. 5 for the sake of clarity. As regards the central portion of Fig. 1, a detail view as shown in Fig. 2 provides a better understanding of the interdigitated structure of the transistor.

The device (see Figs. 1-4) comprises a semiconductor body 1, which is made of silicon in this example, but which may also be made of another suitable semiconductor material, of course. It is provided with an insulating layer 76 of silicon dioxide. The semiconductor body is built up of a low-ohmic, strongly doped p-type substrate 2 and a

5 comparatively weakly doped, high-ohmic region 3 adjoining the surface of the silicon body, in which the transistor is accommodated. In this example, the region 3 is formed by a p-type epitaxial layer having a thickness of approximately 7 μm and a doping concentration of approximately 5.10^{15} atoms per cm^3 . The doping concentration of the substrate 2 which functions as a connection for the source zone is high, for example between 10^{19} and 10^{20}

10 atoms per cm^3 . An active region 6 is defined in the epitaxial layer, which region is laterally bounded by thick field oxide 7. Source and drain zones of the transistor are provided in the active region in the form of strongly doped n-type surface zones 4 and 5, respectively. The transistor comprises a multi-digit structure comprising a number of source/drain digits lying beside one another, which are only shown schematically (Fig. 1) or in part (Fig. 2) in the

15 drawing. The multi-digit structure may be obtained in a simple manner, for example by extending the portion that is shown in Fig. 3 to the left and to the right until the desired channel width is obtained. To increase the breakdown voltage, the drain zone 5 is provided with a high-ohmic n-type drain extension 8 between the drain zone 5 and the channel of the transistor. The length of the extension is 3.5 μm in this example. The transistor channel is

20 formed by the p-type region 13 between the extension 8 and the source zone 4. A gate electrode 9 is provided above the channel, which gate electrode is separated from the channel by a gate oxide 10 having a thickness of, for example, 70 nm. The gate electrode 9 is formed by strips of strongly doped, approximately 0.3 μm thick polycrystalline silicon (poly) overlaid with approximately 0.2 μm titanium silicide, which, seen at the surface, extends

25 transversely over the active region 6 between the source zones 4 and the drain extensions 8. The source zone (or zones) 4 is (are) short-circuited with the p-type region via a deep, strongly doped p-type zone 11 which extends from the surface down to the strongly doped substrate and which connects the source zone 4 to the source electrode 12 at the lower side of the substrate via the substrate 2. The transistor is embodied as an LDMOST, so that it can be

30 operated at a sufficiently high voltage, for which purpose an additional p-type doping is provided in the channel in the form of the diffused p-type zone 13, so that the doping concentration is locally increased as compared with the weak epi doping.

The surface is coated with a thick glass layer, in which contact windows are provided above the source and drain zones, through which windows the source and drain

zones are connected to metal source and drain electrodes 15 and 16, respectively. As is apparent from the plan view of Fig. 2, the contacts 15 and 16 are formed by metal strips extending parallel to each other over the glass layer. The source contact 15 is not only connected to the source zone(s), but also to the deep p-type zone 11, and thus interconnects 5 the source zone and the connection 12 at the bottom side of the substrate. The source zone may be connected to external connections via this connection. The drain electrode strips 16 (Fig. 1) form a comb structure together with the base portion 17, they may be connected to a number of bond pads 116 present elsewhere on the crystal via the common portion 17.

The gate electrode 9 of the device is also provided with a metal contact, which 10 extends in the form of a strip 118 over the oxide layer between the metal strips 15 and 16, and which is locally connected to the gate 9 via contact windows in the oxide layer. The metal track 18 is not connected to the gate 9 over its entire length, but only in a number of interspaced locations, at which the poly gate 9 is provided with widened portions suitable for connections 19, only one of which – for each gate electrode 9 – is shown. If the interspacings 15 between the connections 19 are sufficiently small, the gate resistance is significantly reduced by the presence of the metal tracks 118. The resistance of the gate electrode is also reduced by the presence of titanium silicide thereon. A very low gate resistance can be obtained through the use of a metal having a low resistivity, for example gold or aluminum. As is shown in Fig. 2, the polysilicide (in this example) track of the gate electrode has a comb 20 shape which forms an interdigitated structure with the drain electrode 16, 17. The base of the comb of the gate electrode metal tracks 118 is a metal strip 40, which is provided with evenly spaced bond pads 45 for the gate electrode 9. Adjacent gate electrodes 9 are interconnected, for example in three places – only one of which is shown in Fig. 1 – at the location of 25 connections (not shown in the figure) of the gate electrode 9. In that case a metal track 118, which is connected to the base portion 40 of the gate electrode, extends centrally over said connections. The metal strip 15 of the source region 4 is interrupted at the location of said track 118 so as to reduce the capacitance with the track 118.

Further metal tracks 20 are provided between the polysilicide tracks 18 of the 30 gate electrode 9 and the Al tracks 16 of the drain contact. Said tracks 20 are connected to an electrode 31 of a capacitor 30 and also to a connecting bond pad 35, where an external voltage is applied during operation of the device 10. The (partially interconnected) shielding tracks 20 are connected to the capacitor 30 at evenly spaced positions, said tracks being formed in the lower layer of the two metal layers 20,18 that are separated from each other by means of an insulating silicon dioxide layer 77. The use of a two-metal layer process makes it

possible for the metal tracks 22 to cross the gate electrode 9,18. This makes it possible to connect metal tracks 20 having a minimum resistivity. In this example, another electrode of the capacitor 30 is formed by the portion of the semiconductor body 1 that is present under a thin oxide layer 36, in this case a portion of the epitaxial layer 3 and the substrate 2, which 5 electrode is connected to the source connection 12, therefore. The upper electrode 31 is connected, via metal plugs 34 and an additional metal layer 37 incorporated therein, to a polycrystalline silicon region 99 present on the oxide layer 36 and to the further metal strip 20. In this example, the capacity is 100 pF.

According to the invention, distortion compensation may occur upon 10 presentation of an external voltage to the contact region 35 in the device 10 of this example, owing to the fact that the device comprises two transistors, as it were, an enhancement-type transistor and a depletion-type transistor. As a result, the device exhibits an improved linearity and operation at higher power levels is possible in comparison with the prior art device. It has been found that a suitable external voltage on the metal tracks 20 is about + 25 15 V. All this will be illustrated hereinafter in the discussion of Figs. 6-8.

In a very advantageous variant of the device 10 discussed herein (see Fig. 5), use has been made in the forming of the capacitor 30 of the fact that the device has been formed by means of a two-metal layer process. The lower electrode 32 is formed in the first (lower) metal layer, whilst the upper electrode 31 is formed in the second (upper) metal layer. 20 Present between the two layers is a dielectric of, for example, silicon nitride having a thickness of 100-500 nm. The lower electrode 32 is connected, via metal plugs 34 in an insulating layer, to a silicide region 35 at the surface of the semiconductor body 1, which is further provided with a strongly doped (p-type) semiconductor region 36 at that location. The main advantages of this variant are the fact that a higher capacity value per unit area is 25 possible, and that consequently the surface area of the capacitor 30 may be smaller. The capacity of the capacitor 30 is further substantially independent of the voltage. A final advantage is the fact that the capacitor 30 can easily be connected as a series capacitor, via the lower electrode 32, to a circuit (preferably integrated therewith) capable of applying a power-dependent voltage, for example, to the metal tracks 20. This enables a further 30 improvement and optimization of the distortion and thus use of the device at even higher power levels.

Fig. 6 shows the magnitude S21 as a function of the input power. This complex magnitude S21 is also known as the large signal transducer gain, and it characterizes the linearity of a device, in this case the device of Fig. 1. The large signal gain S21 is defined

as the ratio between the forward wave at the output and the forward wave at the input and can be represented as $r.e^{i\varphi}$. In the figure, the upper group of curves 60 represent the normalized argument (ArgS21) of S21, and the lower group of curves 61 represent the magnitude (Mag 21) of S21 as a function of the input power Pin. The drain-source voltage Vds is 26 V, and 5 the frequency f is 2 GHz. The curves 61a-61e and 60a-60e show the changes in a voltage on the further metal tracks 20, which changes take place in steps (step size is 10 V) from + 20 V to - 20 V. The magnitude MagS21 only depends on the voltage on the metal tracks 20 to a small extent. The magnitude ArgS21, however, exhibits a strongly improved profile (curve 10 60a) for a voltage of + 20 V on the aforesaid metal tracks 20. The (phase), distortion of a device according to the invention - during operation - has been significantly improved in comparison with the prior art device, therefore.

Fig. 7 shows a distortion characterizing factor and the gain of the device of Fig. 1 as a function of the average power. The upper group of curves 70 represent the gain Gp as a function of the average power Pav and the lower group of curves 71 represent the 15 intermodulation distortion IMD3 for a two-tone system, wherein f1 = 2,000 GHz and f2 = 2,001 GHz. Vds is likewise 26 V, and the voltage on the gate electrode 9 is 4.70 V. The curves 70a-70j and 71a-71j represent the changes in a voltage on the further metal tracks 20, which changes take place in steps (step size is 5 V) from + 25 V to - 25 V. The influence of the aforesaid voltage on the gain Gp is small. A significant improvement of IMD3 – at higher 20 power levels – is realized, however, when the aforesaid voltage is approximately + 25 V.

Fig. 8 shows two distortion factors and the efficiency of the device as a function of the average power. A two-carrier WCDMA (= Wideband Code Division Multiple Access) system is concerned here. Curve 80a represents the efficiency (Eff) for a voltage on the metal tracks 20 of +25 V, whilst said voltage is 0 V for the curve 80b. The aforesaid 25 voltage has practically no effect on the efficiency, therefore. The curves 81a and 81b represent the ACPR (= Adjacent Power Ratio) for a metal strip 20 voltage of + 25 V and 0 V, respectively. The curves 82a and 82b represent the magnitude IM3 for corresponding voltages on the further metal strip 20. Fig. 8 shows that the distortion characterizing magnitudes ACPR and IM3 improve significantly at high power levels for a two-carrier 30 system, too, for a voltage on the further metal strip 20 of + 25 V. This figure also illustrates that it may be useful to make the aforesaid voltage dependent on the power level. Thus, the aforesaid voltage may be lower at lower power levels than at higher power levels. The change may take place in steps or continuously, and in both cases it may be applied by a circuit which has been added to the device and which is preferably integrated therewith.

It will be apparent that the invention is not limited to the embodiment described herein, but that many more variations are possible to those skilled in the art within the scope of the invention. Thus the invention may also be used to advantage in p-channel MOS transistors. The invention may furthermore be used not only in DMOS-type transistors, but 5 also in MOS transistors in which the channel is formed by a surface region of the epitaxial layer 3, possibly with additional V_t implantation, instead of a diffused zone. Furthermore it is possible to use the invention in transistors of the so-termed VDMOS type, wherein V stands for vertical, and wherein the channel and the gate electrode still extend (at least substantially) parallel to the surface of the semiconductor body, indicated by the term "semi-lateral" in the 10 present application. Although use in truly vertical MOS transistors, in which the channel and the gate electrode extend substantially perpendicularly to the surface of the semiconductor body, for example in/on the side wall of a groove in said surface, is in itself conceivable, such use is not practically feasible, given the present state of the technology.

Although the device according to the invention preferably comprises a silicon 15 semiconductor body, the invention may also be used in devices made of other semiconductor materials, such as GaAs or other so-termed III-V materials.

It is further noted that while mention is only made of a discrete semiconductor element in the embodiment described herein, the invention is also suitable for the manufacture of other integrated semiconductor products, which may comprise a larger 20 number of active semiconductor elements.

It is further noted that the transistor may be used to great advantage in the 0.5-5 GHz range, in particular in the 0.9-2.4 GHz range, within which range the major frequency bands for mobile communication lie, and that it is suitable for power amplification at lower and also at higher source-drain voltages, such as 20-30 V, for example.